

DERANDOMIZING RESTRICTED ISOMETRIES VIA THE LEGENDRE SYMBOL

AFONSO S. BANDEIRA, MATTHEW FICKUS, DUSTIN G. MIXON, AND JOEL MOREIRA

ABSTRACT. The restricted isometry property (RIP) is an important matrix condition in compressed sensing, but the best matrix constructions to date use randomness. This paper leverages pseudo-random properties of the Legendre symbol to reduce the number of random bits in an RIP matrix with Bernoulli entries. In this regard, the Legendre symbol is not special—our main result naturally generalizes to any small-bias sample space. We also conjecture that no random bits are necessary for our Legendre symbol-based construction.

1. INTRODUCTION

Erdős first used the probabilistic method in 1947 to prove the existence of so-called *Ramsey graphs*, i.e., n -vertex graphs with no clique or independent set of size $O(\log n)$ [21]. Since then, the probabilistic method has been a popular technique for demonstrating the existence of various combinatorial objects, while at the same time leaving a desire for explicit constructions. Even Ramsey graphs currently lack an explicit construction despite the fact that random graphs are (unverifiably) Ramsey with high probability. This aggravatingly common state of affairs has been described by Avi Wigderson as the problem of *finding hay in a haystack*.

Over the last decade, a new combinatorial object has been actively studied in applied mathematics: An $M \times N$ matrix Φ is said to satisfy the (K, δ) -*restricted isometry property (RIP)* if

$$(1 - \delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta)\|x\|_2^2$$

for every vector x with at most K nonzero entries. A matrix which satisfies RIP is well-suited as a *compressive sensor*: If Φ is $(2K, \delta)$ -RIP with $\delta < \sqrt{2} - 1$, then any nearly K -sparse signal x can be stably reconstructed from noisy measurements $y = \Phi x + e$ [12]. Impressively, the number of measurements M required for stable reconstruction is much smaller than the signal dimension N . As we will discuss in the next section, there are several known random constructions of RIP matrices with $M = O_\delta(K \text{ polylog } N)$, meaning the number of measurements can effectively scale with the complexity of the signal.

Unfortunately, RIP matrices provide yet another instance of finding hay in a haystack. Much like Ramsey graphs, RIP matrices are ubiquitous, but verifying RIP is NP-hard in general [5]. To summarize the state of the art, almost every explicit construction of RIP matrices uses the following recipe (see [4, 19, 22], for example):

1. Let the columns $\{\varphi_n\}_{n=1}^N$ of Φ have unit norm and satisfy $|\langle \varphi_n, \varphi_{n'} \rangle| \leq \mu$ for every $n \neq n'$.
2. Observe, either by the Riesz–Thorin theorem or the Gershgorin circle theorem, that every $K \times K$ principal submatrix A of $\Phi^* \Phi$ satisfies $\|A - I\|_2 \leq (K - 1)\mu$.
3. Conclude that Φ is (K, δ) -RIP for $\delta = (K - 1)\mu$.

Date: June 17, 2014.

Key words and phrases. derandomization, Legendre symbol, small-bias sample space, restricted isometry property, compressed sensing.

This coherence-based method of estimating δ requires $K \ll 1/\mu$. However, Welch [38] gives the following universal bound

$$\mu^2 \geq \frac{N - M}{M(N - 1)},$$

which is $\geq 1/(2M)$ whenever $M \leq N/2$. As such, in Vinogradov notation we have

$$M \gg \min\{1/\mu^2, N\} \gg \min\{K^2, N\},$$

meaning a coherence-based guarantee requires M to scale with the *square* of the complexity of the signal. This inferiority to existing guarantees for random constructions compelled Terence Tao to pose the problem of explicitly constructing better RIP matrices on his blog [35]. Since then, one explicit construction has managed to chip away at this power of two—Bourgain, Dilworth, Ford, Konyagin and Kutzarova [9, 10] provided a new recipe, which is largely based on the following higher-order version of coherence:

Definition 1.1. *We say a matrix $\Phi = [\varphi_1 \cdots \varphi_N]$ has (K, θ) -flat restricted orthogonality if*

$$\left| \left\langle \sum_{i \in I} \varphi_i, \sum_{j \in J} \varphi_j \right\rangle \right| \leq \theta(|I||J|)^{1/2}$$

for every disjoint pair of subsets $I, J \subseteq \{1, \dots, N\}$ with $|I|, |J| \leq K$.

Theorem 1.2 (essentially Theorem 13 in [6], cf. Lemma 1 in [9]). *If a matrix Φ with unit-norm columns has (K, θ) -flat restricted orthogonality, then Φ satisfies the $(2K, \delta)$ -restricted isometry property with $\delta = 150\theta \log K$.*

Using this, Bourgain et al. leveraged additive combinatorics to construct an RIP matrix with $M = O_\delta(K^{2-\varepsilon})$ (according to [10], ε is on the order of 10^{-15}); the reader is encouraged to read [29] for a gentle introduction to this construction.

Since explicit constructions are evidently hard to come by, it is prudent to consider a relaxation of the hay-in-a-haystack problem:

Problem 1.3 (Derandomized hay in a haystack). *Given N , K , δ , and a budget H , what is the smallest M for which there exists a random $M \times N$ matrix satisfying the (K, δ) -restricted isometry property with high probability, but whose distribution only uses at most H random bits?*

This derandomization problem is the focus of this paper. The next section features a survey of the current literature. In Section 3, we review our approach, namely, populating the matrix with consecutive Legendre symbols. Our main result (Theorem 3.2) gives that a random seed of $H = O_\delta(K \log K \log N)$ bits suffices to ensure that such a matrix satisfies $(2K, \delta)$ -RIP with high probability; we also conjecture that no random seed is necessary (Conjecture 3.3). Section 4 states a more general version of our main result—that any matrix with small-bias Bernoulli entries satisfies RIP with high probability (Theorem 4.2); this is a natural generalization of a standard result with iid Bernoulli entries, and our main result follows from this one after verifying that random Legendre symbols have small bias. We conclude in Section 5 with the proof of Theorem 4.2.

2. PRIOR WORK ON DERANDOMIZING RESTRICTED ISOMETRIES

When $H = 0$, the explicit construction of Bourgain et al. [9, 10] holds the current record of $M = O_\delta(K^{2-\varepsilon})$, though this construction also requires $N \ll M^{1+\varepsilon}$. If one is willing to spend $H = O_\delta(KN \log(N/K))$ random bits, then simply toss a fair coin for each entry of Φ and populate accordingly with $\pm 1/\sqrt{M}$'s; this matrix is known to be (K, δ) -RIP with high probability provided $M = O_\delta(K \log(N/K))$ [7, 28], and this is the smallest known M for which there exists an $M \times N$ matrix satisfying (K, δ) -RIP. Far less entropy is required if one is willing to accept more structure in Φ . For example, toss N independent weighted coins, each with success probability M_0/N , and

then collect rows of the $N \times N$ discrete Fourier transform matrix which correspond to successful tosses; after scaling the columns to have unit norm, this random partial Fourier matrix will be (K, δ) -RIP with high probability provided $M_0 = O_\delta(K \log^3 K \log N)$; see [13, 14, 33]. The entropy of this construction is $H = O(M_0 \log(N/M_0)) = O_\delta(K \log^3 K \log^2 N)$.

For another construction, this one requiring $N \leq M^2$, toss M fair coins to populate the first column with $\pm 1/\sqrt{M}$'s, and then take the remaining columns to be Gabor translates and modulates of this random seed vector; this matrix is (K, δ) -RIP with high probability provided $M = O_\delta(K \log^2 K \log^2 N)$ [25], and uses entropy $H = M$. In comparison with the random partial Fourier matrix, this construction has slightly more rows in exchange for a log factor of less entropy. Another exchange can be made by modifying the partial Fourier construction: Take a random partial Fourier matrix with BM rows, partition the rows into subcollections of size B , toss B fair coins for each subcollection and negate half of the rows accordingly, and then add the rows in each subcollection to produce a total of M rows; after normalizing the columns, the resulting $M \times N$ matrix will be RIP with high probability provided $B = O(\log^{6.5} N)$ and $M = O_\delta(K \log^2 K \log N)$ [31]. Thanks to B , the entropy in this construction is $H = O_\delta(K \log^2 K \log^{8.5} N)$, and so we save a log factor in M in exchange for several log factors in H .

It should be noted that the preceding random matrix constructions were not developed with the specific intent of derandomization, but rather, they were developed to satisfy certain application-driven structural requirements (or to allow for fast matrix-vector multiplication) and derandomization came as a convenient byproduct. Other constructions have also been proposed with the particular goal of derandomization. For one example, start with a matrix of Bernoulli 0's and 1's, and then for each column, apply an error-correcting binary code to convert the column into a codeword and then change the 0's and 1's to $\pm 1/\sqrt{M}$'s. If the codebook covers $\{0, 1\}^M$ in Hamming distance, then this matrix differs from the corresponding Bernoulli matrix of $\pm 1/\sqrt{M}$'s in only a few entries, resulting in little loss in the δ for RIP. This error-correction trick was introduced in [11], which also identified various codes to derandomize Bernoulli matrices with $M = O_\delta(K \log(N/K))$. Unfortunately, the total entropy for each of these cases is invariably $H = (M - \text{polylog } M)N = \Omega(MN)$; the dual of concatenated simplex codes offers slightly more derandomization $H = (M - M^\varepsilon \log M)N = \Omega(MN)$, but at the price of $M = \Omega(K^{1+\varepsilon})$.

A particularly general RIP construction uses a *Johnson–Lindenstrauss (JL) projection*, namely a random $M \times N$ matrix Φ such that for every $x \in \mathbb{R}^N$ of unit norm,

$$\Pr \left(\left| \|\Phi x\|_2^2 - 1 \right| \geq \varepsilon \right) \leq e^{-cM\varepsilon^2} \quad \forall \varepsilon > 0.$$

JL projections are useful for dimensionality reduction, as they allow one to embed P members of \mathbb{R}^N in $M = O((\log P)/\varepsilon^2)$ dimensions while preserving pairwise distances to within a distortion factor of ε . It was established in [7] that JL projections are (K, δ) -RIP with high probability provided $M = O_\delta(K \log(N/K))$. Moreover, there are several random matrices which are known to be JL, for example, matrices of Bernoulli $\pm 1/\sqrt{M}$'s [1]. Not surprisingly, there has also been work on derandomizing JL projections, though such projections inherently require $H = \Omega_\varepsilon(M)$; since $M < N$, then for every instance Φ_i of the random matrix Φ , there exists a unit-norm vector x in the null space of Φ_i , and so the probability p_i that $\Phi = \Phi_i$ is $\leq e^{-cM\varepsilon^2}$, implying

$$H = \sum_i p_i \log(1/p_i) \geq \min_i \log(1/p_i) \gg_\varepsilon M.$$

This reveals an intrinsic bottleneck in derandomizing JL projections as a means to derandomize RIP matrices. Nevertheless, the most derandomized RIP matrices to date are JL projections.

These particular JL projections exploit r -wise independent Bernoulli random variables, meaning every subcollection of size r is mutually independent. We note that the standard construction of n Bernoulli random variables which are r -wise independent employs an r -wise independent hash family, requires only $O(r \log n)$ random bits, and is computationally efficient. As a consequence of

Theorem 2.2 in [17], a Bernoulli $\pm 1/\sqrt{M}$ matrix with $\Omega_\varepsilon(M)$ -wise independent entries is necessarily JL, thereby producing an RIP matrix with $M = O_\delta(K \log(N/K))$ and $H = O_\delta(K \log(N/K) \log N)$. This has since been derandomized further by Kane and Nelson [24], who multiply a particular sequence of these JL projections to produce a JL family that leads to RIP matrices with $M = O_\delta(K \log(N/K))$ and $H = O_\delta(K \log(N/K) \log(K \log(N/K)))$; at the moment, this is the most derandomized construction with this minimal value of M . In the spirit of derandomizing JL, note that Krahmer and Ward [26] proved that randomly negating columns of an RIP matrix produces a JL projection, establishing that an RIP construction with entropy H leads to a JL family with entropy $H + N$; of course, this is sub-optimal if K is much smaller than N .

3. OUR APPROACH

In this paper, we propose a derandomized version of the matrix of Bernoulli $\pm 1/\sqrt{M}$'s, and our construction makes use of the Legendre symbol. For any integer a and odd prime p , the *Legendre symbol of a and p* is defined to be

$$\left(\frac{a}{p}\right) := \begin{cases} 1 & \text{if } \exists x \in \{1, \dots, p-1\} \text{ such that } a \equiv x^2 \pmod{p} \\ -1 & \text{if } \nexists x \in \{1, \dots, p-1\} \text{ such that } a \equiv x^2 \pmod{p} \\ 0 & \text{if } a \equiv 0 \pmod{p}. \end{cases}$$

The Legendre symbol is known for its pseudorandom behavior, and it is historically associated with derandomization. For example, Erdős used the probabilistic method in [20] to show that almost every way of directing edges in a complete graph forms something called a *paradoxical tournament*, and Graham and Spencer [23] later showed that directing edges between p vertices according to the Legendre symbol (namely, $i \rightarrow j$ precisely when $((i-j)/p) = 1$) provides an explicit construction of such a tournament. We take inspiration from this success in applying the Legendre symbol to find hay in a haystack. Our particular application will make use of the following theorem (historically attributed to Harold Davenport), which establishes higher-order random-like cancellations:

Theorem 3.1 (cf. Theorem 1 in [27]). *There exists p_0 such that for every $p \geq p_0$ and integers $0 < d_1 < \dots < d_k < p$ and $1 \leq t \leq p - d_k$, we have*

$$\left| \sum_{n=0}^{t-1} \left(\frac{n+d_1}{p}\right) \dots \left(\frac{n+d_k}{p}\right) \right| \leq 9kp^{1/2} \log p.$$

If the sum above ranged over all $n \in \mathbb{Z}/p\mathbb{Z}$, then we could invoke Weil's character sum estimate (Theorem 2C in [34]) to get $\leq kp^{1/2}$. Instead, the sum is truncated so as to avoid a zero Legendre symbol, and so we are penalized with a log factor—this estimate is made possible by a version of the Erdős–Turán inequality.

For one interpretation of this result, draw X uniformly from $\{0, \dots, p-k-1\}$. Then all moments and mixed moments of the random Legendre symbols $\{((X+i)/p)\}_{i=1}^k$ have size $\leq (9kp^{1/2} \log p)/(p-k-1)$, which vanishes if $k = o(p^{1/2}/\log p)$. As such, these random Legendre symbols have low correlation, but notice how little entropy we used. Indeed, at the price of only $H < \log_2 p$ random bits to pick X , we can enjoy as many as, say, $k = p^{1/3}$ bits of low correlation. This quasi-paradox reveals something subtle about Bernoulli random variables: While zero correlation implies independence (meaning the total entropy is the number of random variables), it is possible to produce many Bernoulli random variables of low correlation with very little entropy; as we discuss in the next section, this is related to packing many nearly orthogonal vectors in low-dimensional Euclidean space.

In this paper, we leverage the pseudorandomness identified in Theorem 3.1 to derandomize the matrix of Bernoulli $\pm 1/\sqrt{M}$'s while maintaining RIP. Our particular matrix construction uses the following recipe:

1. Given N, K and δ , pick M and H sufficiently large, and take p to be some prime $\geq 2^H + MN$.

2. Draw X uniformly from $\{0, \dots, 2^H - 1\}$, and populate the entries of an $M \times N$ matrix Φ one column at a time with consecutive Legendre symbols $\{((X + i)/p)\}_{i=1}^{MN}$.
3. Use Theorem 3.1 to show that Φ has flat restricted orthogonality, and therefore satisfies (K, δ) -RIP by Theorem 1.2.

The reader should rest assured that the third step above is not obvious, but it summarizes our proof technique for the main result of this paper:

Theorem 3.2 (Main Result). *Given N , K and δ , take*

$$M = (C_1/\delta^2)K \log^2 K \log N, \quad H = C_2 K \log((K/\delta) \log K) \log N$$

with C_1 and C_2 sufficiently large, and let p denote any prime $\geq 2^H + MN$. Draw X uniformly from $\{0, \dots, 2^H - 1\}$, and define the corresponding $M \times N$ matrix Φ entrywise by the Legendre symbol

$$\Phi[m, n] := \frac{1}{\sqrt{M}} \left(\frac{X + M(n-1) + m}{p} \right).$$

Then Φ satisfies the $(2K, \delta)$ -restricted isometry property with high probability.

At this point, we explain how to efficiently construct this matrix. Note that the output (i.e., the $M \times N$ matrix Φ) is of length $MN \leq N^2$, and so consider a construction algorithm to be computationally efficient if it takes time polynomial in N . The first task is to find a prime p which is $\geq 2^H + MN$. By Bertrand's postulate, it suffices to search the integers between $2^H + MN$ and $2(2^H + MN)$, of which a fraction of about $1/\log(2^H + MN) = \Omega(1/H)$ are primes by the prime number theorem. As such, we may randomly draw an integer in this interval, run the AKS primality test [2], and repeat until we find a prime; this randomized algorithm will succeed in $\text{poly}(H) = \text{poly}(N)$ time with high probability. (We note that deterministic alternatives to this approach have been studied, for example, in a polymath project [36], but the best known algorithms of this form use superpolynomial time.) Once the prime p has been selected, we draw X uniformly from $\{0, \dots, 2^H - 1\}$, which requires $H = O(N)$ queries of a Bernoulli random number generator. Finally, the entries of Φ are populated with $MN \leq N^2$ different Legendre symbols, and each of these can be calculated in $\text{polylog}(p) = \text{poly}(N)$ time, either by appealing to Euler's criterion (using repeated squaring with intermediate reductions modulo p) or by exploiting various identities of the Jacobi symbol (which equals the Legendre symbol whenever the “denominator” is prime) [18].

We now identify the shortcomings of our main result. First, the number of measurements M that scales like $K \log^2 K \log N$ instead of the minimal $K \log(N/K)$. We credit this (admittedly small) difference to our use of flat restricted orthogonality (i.e., Theorem 1.2) to demonstrate RIP; indeed, Theorem 14 in [6] gives that using flat restricted orthogonality to prove RIP of Gaussian matrices, which are known to be RIP with minimal $M = O_\delta(K \log(N/K))$ [13], leads to a scaling in M that is identical to Theorem 3.2, and so our result can perhaps be strengthened by somehow proving RIP “directly.” Next, we note that our construction uses entropy H that scales like $K \log K \log N$, which is slightly more than the JL construction of Kane and Nelson [24] that uses $O_\delta(K \log(N/K) \log(K \log(N/K)))$ random bits. As we noted earlier, JL constructions of RIP matrices necessarily use at least $\Omega_\delta(M) = \Omega_\delta(K \log(N/K))$ random bits, whereas we believe our Legendre symbol construction can be derandomized quite a bit more:

Conjecture 3.3. *There exists a universal constant C such that for every $\delta > 0$, there exists $N_0 > 0$ and $P_0(N) = O(2^{\text{poly}(N)})$ such that for every (K, M, N) satisfying*

$$M \geq (C/\delta^2)K \log(N/K), \quad N \geq N_0,$$

and for every prime $p \geq P_0$, the $M \times N$ matrix Φ defined entrywise by the Legendre symbol

$$\Phi[m, n] := \frac{1}{\sqrt{M}} \left(\frac{M(n-1) + m}{p} \right)$$

satisfies the $(2K, \delta)$ -restricted isometry property.

At its heart, Conjecture 3.3 is a statement about how well the Legendre symbol exhibits additive cancellations. In particular, if one is willing to use flat restricted orthogonality as a proof technique for RIP, the statement is essentially a bound on incomplete sums of Legendre symbols, much like those investigated in [16]. We had difficulty proving these cancellations, and so we injected some randomness to enable the application of Theorem 3.1. Since Conjecture 3.3 could very well be beyond the reach of modern techniques, we pose the following weaker and more general problem, whose solution would be a substantial advance in the derandomization of RIP (since this level of derandomization is not achievable with JL projections):

Problem 3.4 (Breaking the Johnson–Lindenstrauss bottleneck). *Find a construction of $M \times N$ matrices which satisfies the (K, δ) -restricted isometry property with high probability whenever $M = O_\delta(K \text{ polylog } N)$ and uses only $H = o_{\delta; N \rightarrow \infty}(K \log(N/K))$ random bits.*

4. THE MAIN RESULT

In this section, we reduce our main result to a more general (and independently interesting) result, which requires the following definition:

Definition 4.1. *Let $\{X_i\}_{i=1}^n$ be a sequence of Bernoulli random variables taking values in $\{\pm 1\}$. We say $\{X_i\}_{i=1}^n$ is ε -biased if*

$$\left| \mathbb{E} \prod_{i \in I} X_i \right| \leq \varepsilon \quad (1)$$

for every nonempty $I \subseteq \{1, \dots, n\}$.

This notion of bias was introduced by Vazirani in [37]. We will use bias as a means for derandomizing RIP matrices. In particular, we will show that instead of using iid Bernoulli entries, it suffices to have small-bias Bernoulli entries for RIP, and then we will apply Theorem 3.1 to verify that random Legendre symbols have sufficiently small bias (even though the required entropy is small). The idea that small-bias random variables can be used for derandomization is not new; indeed, one of their primary applications is to reduce the number of random bits needed for randomized algorithms [30]. As promised, the following generalizes the RIP result for matrices with independent Bernoulli entries to matrices with small-bias Bernoulli entries:

Theorem 4.2. *Fix $C = 5760000$. There exists N_0 for which the following holds: Given $N \geq N_0$, $K \geq 1$ and $\delta > 0$, take $M = (C/\delta^2)K \log^2 K \log N$ and pick*

$$\varepsilon \leq \exp \left(-40K \log((150/\delta)K \log K) \log N \right).$$

Then any $M \times N$ matrix populated with ε -biased Bernoulli random variables (scaled by $1/\sqrt{M}$) satisfies the $(2K, \delta)$ -restricted isometry property with probability $\geq 1 - 2N^{-2K}$.

We note that the constants in this theorem have not been optimized. Also, taking $\varepsilon = 0$ recovers the iid Bernoulli result, though with extra log factors appearing in M . We will prove this result in the next section. For now, we establish that Theorem 4.2 implies our main result (Theorem 3.2). Drawing X uniformly from $\{0, \dots, 2^H - 1\}$, we bound the bias of the Legendre symbols $\{((X + i)/p)\}_{i=1}^{MN}$ using Theorem 3.1:

$$\left| \mathbb{E} \prod_{i \in I} \left(\frac{X + i}{p} \right) \right| = \left| \frac{1}{2^H} \sum_{x=0}^{2^H-1} \prod_{i \in I} \left(\frac{x + i}{p} \right) \right| \leq \frac{1}{2^H} |I| p^{1/2} \log p.$$

Since $|I| \leq MN$ and further $p \leq 2(2^H + MN) \leq 4 \cdot 2^H$ by Bertrand's postulate, we have

$$\frac{1}{2^H} |I| p^{1/2} \log p \leq \frac{1}{2^H} MN (4 \cdot 2^H)^{1/2} (\log 4 + H \log 2) \leq 4MNH \cdot 2^{-H/2} \leq 4N^2 \cdot 2^{-H/3}.$$

As such, the random Legendre symbols we use to populate Φ are $(4N^2 \cdot 2^{-H/3})$ -biased. To use Theorem 4.2, it remains to verify that this bias is sufficiently small:

$$\log 4 + 2 \log N - \frac{H}{3} \log 2 \leq -40K \log((150/\delta)K \log K) \log N,$$

which is indeed satisfied by our choice of H (i.e., taking C_2 to be sufficiently large).

Of course, we are not the first to use the Legendre symbol to produce small-bias random variables (for example, see [3, 32]). Also, there are several other constructions of small-bias random variables (e.g., [3, 8, 30]), but these do not naturally lead to a conjecture such as Conjecture 3.3. Interestingly, a stronger version of the Chowla conjecture [15] implies that a randomly seeded portion of the Liouville function also produces small-bias random variables, and so one might pose a corresponding version of Conjecture 3.3. Unfortunately, a large class of small-bias random variables cannot be used in conjunction with Theorem 4.2 to break the Johnson–Lindenstrauss bottleneck (Problem 3.4). To see this, we first make an identification between small-bias random variables and linear codes.

A linear code $\mathcal{C} \subseteq \mathbb{F}_2^q$ is a subspace of *codewords*. The $n \times q$ generator matrix G of an n -dimensional linear code \mathcal{C} has the property that $\mathcal{C} = \{xG : x \in \mathbb{F}_2^n\}$, i.e., the rows of G form a basis for \mathcal{C} . The *weight* of a codeword is the number of entries with value 1.

Proposition 4.3 (cf. [3, 8, 30]).

- (a) Let G be the generator matrix of an n -dimensional linear code in \mathbb{F}_2^q such that every nonzero codeword has weight between $(1 - \varepsilon)q/2$ and $(1 + \varepsilon)q/2$. Randomly sample j uniformly over $\{1, \dots, q\}$ and take $X_i := (-1)^{G_{ij}}$ for every $i = 1, \dots, n$. Then $\{X_i\}_{i=1}^n$ is ε -biased.
- (b) Let $\{X_i\}_{i=1}^n$ be an ε -biased sequence drawn uniformly from some multiset $\mathcal{X} \subseteq \{\pm 1\}^n$ of size q (\mathcal{X} is called an ε -biased set). For each $x \in \mathcal{X}$, consider the corresponding $g \in \mathbb{F}_2^n$ such that $x_i = (-1)^{g_i}$ for every $i = 1, \dots, n$. Then the matrix G whose columns are the g 's corresponding to x 's in \mathcal{X} is the generator matrix of an n -dimensional linear code in \mathbb{F}_2^q such that every nonzero codeword has weight between $(1 - \varepsilon)q/2$ and $(1 + \varepsilon)q/2$.

To see the significance of this proposition, consider the n -dimensional linear code $\mathcal{C} \subseteq \mathbb{F}_2^q$ corresponding to an ε -biased set of size q . For each codeword $c \in \mathcal{C}$, define a unit vector $v_c \in \mathbb{R}^q$ whose entries have the form $(-1)^{c_i}/\sqrt{q}$. It is easy to verify that $|\langle v_c, v_{c'} \rangle| \leq \varepsilon$ whenever $c \neq c'$, and so the Welch bound [38] gives

$$\varepsilon^2 \geq \frac{2^n - q}{q(2^n - 1)}.$$

For our application, we have $n = MN$ and $q = 2^H$. If $H \leq MN - 1$ (i.e., $q \leq 2^{n-1}$), then the Welch bound implies $\varepsilon^2 \geq 1/(2q) = 2^{-(H+1)}$. This gives the following result:

Proposition 4.4. Every ε -biased set $\mathcal{X} \subseteq \{\pm 1\}^n$ has entropy $H \geq \min\{\log(1/\varepsilon), n - 1\}$.

As such, applying Theorem 4.2 with any ε -biased set requires $H \gg_\delta K \log K \log N$. In this sense, the Legendre construction in Theorem 3.2 is optimal. On the other hand, no ε -biased set can be used with Theorem 4.2 to break the Johnson–Lindenstrauss bottleneck.

5. PROOF OF THEOREM 4.2

We will first show that Φ has (K, θ) -flat restricted orthogonality, and then appeal to Theorem 1.2 to get the $(2K, \delta)$ -restricted isometry property. To this end, fix a disjoint pair of subsets $I, J \subseteq \{1, \dots, N\}$ with $|I|, |J| \leq K$. We seek to bound the following probability:

$$\Pr \left[\left| \left\langle \sum_{i \in I} \varphi_i, \sum_{j \in J} \varphi_j \right\rangle \right| > \theta(|I||J|)^{1/2} \right] = \Pr \left[\left| \sum_{i \in I} \sum_{j \in J} \sum_{m=1}^M \Phi[m, i] \Phi[m, j] \right| > \theta(|I||J|)^{1/2} \right] \quad (2)$$

Applying a version of Markov's inequality then gives

$$(2) \leq \frac{1}{(\theta(|I||J|)^{1/2})^q} \mathbb{E} \left[\left(\sum_{i \in I} \sum_{j \in J} \sum_{m=1}^M \Phi[m, i] \Phi[m, j] \right)^q \right], \quad (3)$$

for some even integer q which we will optimize later. Observe that, because q is even, the absolute value inside the expectation was not needed. For now, we expand the product of sums and use linearity of expectation to get

$$(3) = \frac{1}{(\theta(|I||J|)^{1/2})^q} \sum_{(i_1, j_1, m_1) \in I \times J \times [M]} \cdots \sum_{(i_q, j_q, m_q) \in I \times J \times [M]} \mathbb{E} \left[\prod_{v=1}^q \Phi[m_v, i_v] \Phi[m_v, j_v] \right].$$

Next, since (3) is nonnegative, the triangle inequality gives

$$(3) = |(3)| \leq \frac{1}{(\theta(|I||J|)^{1/2})^q} \sum_{\{(i_v, j_v, m_v)\}_{v=1}^q \in (I \times J \times [M])^q} \left| \mathbb{E} \left[\prod_{v=1}^q \Phi[m_v, i_v] \Phi[m_v, j_v] \right] \right|.$$

Most of the terms in the sum take the form of (1), and so we can bound their contributions accordingly. Each of the remaining terms has the property that all of its factors appear an even number of times in the product, and we will control the contribution of these terms by establishing how few they are. With this in mind, we now define an index subset $S \subseteq (I \times J \times [M])^q$ (think “surviving” indices). Explicitly, $\{(i_v, j_v, m_v)\}_{v=1}^q \in S$ if for every v , each of the sets

$$\{v' \in [q] : (i_{v'}, m_{v'}) = (i_v, m_v)\}, \quad \{v' \in [q] : (j_{v'}, m_{v'}) = (j_v, m_v)\}$$

has an even number of elements.

We are able to use ε -bias to control all of the terms not in S . Indeed, let $\{(i_v, j_v, m_v)\}_{v=1}^q \in S^c$. Then there are $t \geq 1$ distinct entries of Φ that appear in the product in (3) an odd number of times (observe that I and J are disjoint and hence an entry can not be simultaneously of the form $\Phi[m_v, i_v]$ and $\Phi[m_v, j_v]$). Since there are a total of $2q$ entries of Φ in the product and the square of any entry is $1/M$, we can apply ε -bias to deduce that

$$\left| \mathbb{E} \left[\prod_{v=1}^q \Phi[m_v, i_v] \Phi[m_v, j_v] \right] \right| \leq M^{t/2-q} \frac{\varepsilon}{M^{t/2}} = M^{-q} \varepsilon. \quad (4)$$

Overall, the sum in our bound becomes

$$\begin{aligned} & \sum_{\{(i_v, j_v, m_v)\}_{v=1}^q \in S} \left| \mathbb{E} \left[\prod_{v=1}^q \Phi[m_v, i_v] \Phi[m_v, j_v] \right] \right| + \sum_{\{(i_v, j_v, m_v)\}_{v=1}^q \in S^c} \left| \mathbb{E} \left[\prod_{v=1}^q \Phi[m_v, i_v] \Phi[m_v, j_v] \right] \right| \\ & \leq M^{-q} (|S| + \varepsilon |S^c|), \end{aligned}$$

where the inequality applies the fact that $|\Phi[m, n]| = 1/\sqrt{M}$ to each term from S and (4) to each term from S^c . At this point, we wish to bound $|S|$.

For each $s = \{(i_v, j_v, m_v)\}_{v=1}^q \in S$, there exist two perfect matchings of $\{1, \dots, q\}$, say \mathcal{M}_1 and \mathcal{M}_2 , such that

$$(i_v, m_v) = (i_{v'}, m_{v'}) \quad \forall \{v, v'\} \in \mathcal{M}_1, \quad (j_v, m_v) = (j_{v'}, m_{v'}) \quad \forall \{v, v'\} \in \mathcal{M}_2. \quad (P[s, \mathcal{M}_1, \mathcal{M}_2])$$

(Here, we use the phrase “perfect matching” so as to convey that $\{1, \dots, q\}$ is partitioned into sets of size 2. Let $\mathcal{M}(q)$ denote all such perfect matchings.) We now start our bound:

$$\begin{aligned} |S| &= \#\left\{s \in (I \times J \times [M])^q : \exists \mathcal{M}_1, \mathcal{M}_2 \in \mathcal{M}(q) \text{ such that } P[s, \mathcal{M}_1, \mathcal{M}_2]\right\} \\ &\leq \#\left\{(s, \mathcal{M}_1, \mathcal{M}_2) : P[s, \mathcal{M}_1, \mathcal{M}_2]\right\} \\ &= \sum_{\mathcal{M}_1 \in \mathcal{M}(q)} \sum_{\mathcal{M}_2 \in \mathcal{M}(q)} \#\left\{s : P[s, \mathcal{M}_1, \mathcal{M}_2]\right\}. \end{aligned}$$

Next, given a perfect matching $\mathcal{M} \in \mathcal{M}(q)$, let $C(\mathcal{M})$ denote the set of q -tuples $\{m_v\}_{v=1}^q$ satisfying $m_v = m_{v'}$ whenever $\{v, v'\} \in \mathcal{M}$ (these are the q -tuples which consistently “color” the matching). Note that the q -tuple of m_v ’s from $s = \{(i_v, j_v, m_v)\}_{v=1}^q$ satisfying $P[s, \mathcal{M}_1, \mathcal{M}_2]$ lies in both $C(\mathcal{M}_1)$ and $C(\mathcal{M}_2)$. Moreover, the q -tuple of i_v ’s from $s = \{(i_v, j_v, m_v)\}_{v=1}^q$ satisfying $P[s, \mathcal{M}_1, \mathcal{M}_2]$ assigns members of I to the $q/2$ pairs $\{v, v'\} \in \mathcal{M}_1$, meaning the q -tuple is one of $|I|^{q/2}$ such possibilities; similarly, there are $|J|^{q/2}$ possibilities for the q -tuple of j_v ’s. As such, we can continue our bound:

$$|S| \leq |I|^{q/2} |J|^{q/2} \sum_{\mathcal{M}_1 \in \mathcal{M}(q)} \sum_{\mathcal{M}_2 \in \mathcal{M}(q)} |C(\mathcal{M}_1) \cap C(\mathcal{M}_2)|.$$

At this point, we note that $\sum_{\mathcal{M}_2 \in \mathcal{M}(q)} |C(\mathcal{M}_1) \cap C(\mathcal{M}_2)|$ does not depend on \mathcal{M}_1 , since any permutation of $\{1, \dots, q\}$ that sends one version of \mathcal{M}_1 to another can be viewed as a permutation acting on \mathcal{M}_2 (thereby merely permuting the summands). As such, we can express our bound in terms of a “canonical matching” $\mathcal{M}_0 := \{\{1, 2\}, \{3, 4\}, \dots, \{q-1, q\}\}$:

$$|S| \leq |I|^{q/2} |J|^{q/2} |\mathcal{M}(q)| \sum_{\mathcal{M}_2 \in \mathcal{M}(q)} |C(\mathcal{M}_0) \cap C(\mathcal{M}_2)|.$$

Note that $|\mathcal{M}(q)| = (q-1)!!$ since there are $q-1$ possible members to match with 1, and then $q-3$ members remaining to match with the smallest free element, etc. It remains to determine the sum over $\mathcal{M}(q)$. To parse this sum, consider a graph of q vertices with edge set $\mathcal{M}_0 \cup \mathcal{M}_2$. Then $|C(\mathcal{M}_0) \cap C(\mathcal{M}_2)|$ counts the number of ways of coloring the vertices using M colors in such a way that each component (which is necessarily a cycle) has vertices of a common color. We will use this interpretation to prove the following claim:

$$\sum_{\mathcal{M}_2 \in \mathcal{M}(q)} |C(\mathcal{M}_0) \cap C(\mathcal{M}_2)| = \frac{(M+q-2)!!}{(M-2)!!}.$$

To proceed, start with a graph G of q vertices, and let the edge set be \mathcal{M}_0 . Let A denote the “available” set of vertices, and initialize A as $\{1, \dots, q\}$. We will iteratively update the pair (G, A) as we build all possible matchings \mathcal{M}_2 and colorings. At each step, denote $k := |A|$. Take the least-numbered available vertex and draw an edge between it and any of the other $k-1$ available vertices. Add this edge to \mathcal{M}_2 and remove both vertices from A . If this edge completes a cycle in G , then color that cycle with any of the M colors. Overall, in this step you either pick one of $k-2$ vertices which do not complete the cycle, or you complete the cycle and pick one of M colors; this totals to $M+k-2$ choices at each step. Since the number k of available vertices decreases by 2 at each step, the product of these totals corresponds to the claim.

At this point, we summarize the status of our bound:

$$\begin{aligned} \Pr \left[\left| \left\langle \sum_{i \in I} \varphi_i, \sum_{j \in J} \varphi_j \right\rangle \right| > \theta(|I||J|)^{1/2} \right] &\leq \frac{1}{(M\theta(|I||J|)^{1/2})^q} \left(|S| + \varepsilon |S^c| \right) \\ &\leq \frac{(q-1)!!(M+q-2)!!}{(M\theta)^q (M-2)!!} + \frac{\varepsilon |I|^{q/2} |J|^{q/2}}{\theta^q}, \end{aligned} \quad (5)$$

where the last inequality applies our bound on $|S|$ along with $|S^c| \leq |(I \times J \times [M])^q| = (|I||J|M)^q$. To continue, we will bound the terms in (5) separately. When M is even (otherwise, $M-1$ is even and similar bounds will apply), the first term has the following bound:

$$\begin{aligned} \frac{(q-1)!!(M+q-2)!!}{(M\theta)^q(M-2)!!} &\leq \frac{q!!(M+q-2)!!}{(M\theta)^q(M-2)!!} \\ &\leq \frac{2^q((\frac{q}{2})!)^2}{(M\theta)^q} \left(\frac{M}{2} + \frac{q}{2} - 1 \right) \leq \frac{2^q e^2 (\frac{q/2+1}{e})^{q+2}}{(M\theta)^q} \left(e \cdot \frac{M}{2} + \frac{q}{2} - 1 \right)^{q/2}, \end{aligned}$$

where the last step uses Stirling-type bounds on both the factorial and the binomial coefficient. Next, we combine like exponents to get

$$\begin{aligned} \frac{(q-1)!!(M+q-2)!!}{(M\theta)^q(M-2)!!} &\leq e^2 \left(\frac{q/2+1}{e} \right)^2 \left(\frac{4}{(M\theta)^2} \left(\frac{q/2+1}{e} \right)^2 \cdot e \cdot \left(\frac{M}{q} + 1 - \frac{2}{q} \right) \right)^{q/2} \\ &\leq q^2 \left(\frac{4}{(M\theta)^2} \left(\frac{q}{e} \right)^2 \cdot e \cdot \frac{2M}{q} \right)^{q/2}, \end{aligned}$$

where the last step assumes $q \geq 2$ (given our choice of q later, this will occur for sufficiently large N) so that $q/2+1 \leq q$ and also assumes $q \leq M$ (which will follow from the fact that $\delta \leq 1$) so that $M/q+1 \leq 2M/q$. At this point, we take $q = M\theta^2/8$ (or more precisely, take q to be the smallest even integer larger than $M\theta^2/8$) to get

$$\frac{(q-1)!!(M+q-2)!!}{(M\theta)^q(M-2)!!} \leq \left(\frac{M\theta^2}{8} \right)^2 \cdot e^{-M\theta^2/16}.$$

This can be simplified by noting that $M = (C/\delta^2)K \log^2 K \log N$ and $\theta = \delta/(150 \log K)$ together give $M\theta^2 = 256K \log N$:

$$\frac{(q-1)!!(M+q-2)!!}{(M\theta)^q(M-2)!!} \leq \exp \left(2 \log \left(\frac{M\theta^2}{8} \right) - \frac{M\theta^2}{16} \right) \leq \exp \left(-8K \log N \right), \quad (6)$$

for sufficiently large N . Next, we bound the second term in (5):

$$\frac{\varepsilon |I|^{q/2} |J|^{q/2}}{\theta^q} \leq \varepsilon (K/\theta)^q = \varepsilon ((150/\delta)K \log K)^q \leq \exp \left(-8K \log((150/\delta)K \log K) \log N \right) \quad (7)$$

Now we can combine the bounds (6) and (7) to get

$$\begin{aligned} \Pr \left[\left| \left\langle \sum_{i \in I} \varphi_i, \sum_{j \in J} \varphi_j \right\rangle \right| > \theta (|I||J|)^{1/2} \right] \\ \leq \exp \left(-8K \log N \right) + \exp \left(-8K \log((150/\delta)K \log K) \log N \right) \\ \leq 2 \exp \left(-8K \log N \right). \end{aligned}$$

At this point, we are ready to show that Φ has the $(2K, \delta)$ -restricted isometry property with high probability. This calculation will use a union bound over all possible choices of disjoint $I, J \subseteq \{1, \dots, N\}$ with $|I|, |J| \leq K$, of which there are

$$\sum_{|I|=1}^K \sum_{|J|=1}^K \binom{N}{|I|} \binom{N-|I|}{|J|} \leq K^2 \binom{N}{K}^2 \leq K^2 \left(\frac{eN}{K} \right)^{2K} \leq \exp \left(4K + 2K \log(N/K) \right).$$

With this, we apply a union bound to get

$$\begin{aligned}
\Pr \left[\Phi \text{ is not } (2K, \delta)\text{-RIP} \right] &\leq \Pr \left[\Phi \text{ does not have } (K, \theta)\text{-FRO} \right] \\
&\leq \exp \left(4K + 2K \log(N/K) \right) \cdot \Pr \left[\left| \left\langle \sum_{i \in I} \varphi_i, \sum_{j \in J} \varphi_j \right\rangle \right| > \theta(|I||J|)^{1/2} \right] \\
&\leq 2 \exp \left(4K + 2K \log(N/K) - 8K \log N \right) \\
&\leq 2 \exp \left(-2K \log N \right),
\end{aligned}$$

which completes the result.

ACKNOWLEDGMENTS

The authors thank Prof. Peter Sarnak for insightful discussions. A. S. Bandeira was supported by AFOSR Grant No. FA9550-12-1-0317, and M. Fickus and D. G. Mixon were supported by NSF Grant No. DMS-1321779. The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

REFERENCES

- [1] D. Achlioptas, Database-friendly random projections: Johnson-Lindenstrauss with binary coins, *J. Comput. Syst. Sci.* 66 (2003) 671–687.
- [2] M. Agrawal, N. Kayal, N. Saxena, PRIMES is in P, *Ann. Math.* 160 (2004) 781–793.
- [3] N. Alon, O. Goldreich, J. Hästad, R. Paralta, Simple constructions of almost k-wise independent random variables, *Random Struct. Algor.* 3 (1992) 289–304.
- [4] L. Applebaum, S. D. Howard, S. Searle, R. Calderbank, Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery, *Appl. Comput. Harmon. Anal.* 26 (2009) 283–290.
- [5] A. S. Bandeira, E. Dobriban, D. G. Mixon, W. F. Sawin, Certifying the restricted isometry property is hard, *IEEE Trans. Inform. Theory* 59 (2013) 3448–3450.
- [6] A. S. Bandeira, M. Fickus, D. G. Mixon, P. Wong, The road to deterministic matrices with the restricted isometry property, *J. Fourier Anal. Appl.* 19 (2013) 1123–1149.
- [7] R. G. Baraniuk, M. Davenport, R. A. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices, *Constr. Approx.* 28 (2008) 253–263.
- [8] A. Ben-Aroya, A. Ta-Shma, Constructing small-bias sets from algebraic-geometric codes, *FOCS 2009*, 191–197.
- [9] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, D. Kutzarova, Explicit constructions of RIP matrices and related problems, *Duke Math. J.* 159 (2011) 145–185.
- [10] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, D. Kutzarova, Breaking the k^2 barrier for explicit RIP matrices, *STOC 2011*, 637–644.
- [11] R. Calderbank, S. Jafarpour, M. Nastasescu, Covering radius and the restricted isometry property, *IEEE Inform. Theory Workshop* (2011) 558–562.
- [12] E. J. Candès, The restricted isometry property and its implications for compressed sensing, *C. R. Acad. Sci. Paris, Ser. I* 346 (2008) 589–592.
- [13] E. J. Candès, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies?, *IEEE Trans. Inform. Theory* 52 (2006) 5406–5425.
- [14] M. Cheraghchi, V. Guruswami, A. Velingker, Restricted isometry of Fourier matrices and list decodability of random linear codes, *SIAM J. Comput.* 42 (2013) 1888–1914.
- [15] S. Chowla, The Riemann hypothesis and Hilbert’s tenth problem, *Mathematics and its applications*, vol. 4, 1965.
- [16] F. R. K. Chung, Several generalizations of Weil sums, *J. Number Theory* 49 (1994) 95–106.
- [17] K. L. Clarkson, D. P. Woodruff, Numerical linear algebra in the streaming model, *STOC 2009*, 205–214.
- [18] H. Cohen, *A Course in Computational Algebraic Number Theory*, Springer, 1993.
- [19] R. A. DeVore, Deterministic constructions of compressed sensing matrices, *J. Complexity* 23 (2007) 918–925.
- [20] P. Erdős, On a problem in graph theory, *Math. Gaz.* 47 (1963) 220–223.
- [21] P. Erdős, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* 53 (1947) 292–294.

- [22] M. Fickus, D. G. Mixon, J. C. Tremain, Steiner equiangular tight frames, *Linear Algebra Appl.* 436 (2012) 1014–1027.
- [23] R. L. Graham, J. H. Spencer, A constructive solution to a tournament problem, *Canad. Math. Bull.* 14 (1971) 45–48.
- [24] D. M. Kane, J. Nelson, A derandomized sparse Johnson-Lindenstrauss transform, Available online: [arXiv:1006.3585](https://arxiv.org/abs/1006.3585)
- [25] F. Krahmer, S. Mendelson, H. Rauhut, Suprema of chaos processes and the restricted isometry property, *Comm. Pure Appl. Math.*, to appear.
- [26] F. Krahmer, R. Ward, New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property, *SIAM J. Math. Anal.* 43 (2011) 1269–1281.
- [27] C. Mauduit, A. Sárközy, On finite pseudorandom binary sequences I: Measure of pseudorandomness, the Legendre symbol, *Acta Arith.* 82 (1997) 365–377.
- [28] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Uniform uncertainty principle for Bernoulli and subgaussian ensembles, *Constr. Approx.* 28 (2009) 277–289.
- [29] D. G. Mixon, Explicit matrices with the restricted isometry property: Breaking the square-root bottleneck, Available online: [arXiv:1403.3427](https://arxiv.org/abs/1403.3427)
- [30] J. Naor, M. Naor, Small-bias probability spaces: efficient constructions and applications, *STOC 1990*, 213–223.
- [31] J. Nelson, E. Price, M. Wootters, New constructions of RIP matrices with fast multiplication and fewer rows, *SODA 2014*, 1515–1528.
- [32] R. Peralta, On the randomness complexity of algorithms, U. Wisconsin, Milwaukee, CS Research Report TR 90-1.
- [33] M. Rudelson, R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, *Comm. Pure Appl. Math.* 61 (2008) 1025–1045.
- [34] W. M. Schmidt, *Equations over finite fields: An elementary approach*, Lecture Notes in Mathematics, Springer Verlag, 1976.
- [35] T. Tao, Open question: deterministic UUP matrices, <http://terrytao.wordpress.com/2007/07/02/open-question-deterministic-uup-matrices/>
- [36] T. Tao, E. Croot III, H. Helfgott, Deterministic methods to find primes, *Math. Comput.* 81 (2012) 1233–1246.
- [37] U. Vazirani, *Randomness, adversaries and computation*, Ph.D. Thesis, U. California, Berkeley, 1986.
- [38] L. R. Welch, Lower bounds on the maximum cross correlation of signals, *IEEE Trans. Inform. Theory* 20 (1974) 397–399.

(Bandeira) PROGRAM IN APPLIED AND COMPUTATIONAL MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NJ 08544, USA (ajsb@math.princeton.edu).

(Fickus) DEPARTMENT OF MATHEMATICS AND STATISTICS, AIR FORCE INSTITUTE OF TECHNOLOGY, WRIGHT-PATTERSON AFB, OH 45433, USA (matthew.fickus@afit.edu).

(Mixon) DEPARTMENT OF MATHEMATICS AND STATISTICS, AIR FORCE INSTITUTE OF TECHNOLOGY, WRIGHT-PATTERSON AFB, OH 45433, USA (dustin.mixon@afit.edu).

(Moreira) DEPARTMENT OF MATHEMATICS, OHIO STATE UNIVERSITY, COLUMBUS, OH 43210, USA (moreira.6@osu.edu).